Angela Hamilton

Career Foundry Project 6

Sourcing Open Data

**Chocolate Bar Ratings**

**Data Sourcing:**  This is an external data source from Kaggle, a subsidiary of Google. The data source is an online platform for data scientists and machine learning enthusiasts, which provides a wide collection of datasets that are available to the public.

**Data Collection**: Open-Source data that is free and available to the public. Since this dataset is considered open data that is collected by third-party sources, there could be bias in the data collection. The last update for this dataset was back in 2017, which causes there to be a time lag.

**Data Content**:

The data contains expert ratings for the different flavors of cocoa of over 1,700 chocolate bars from 2006 – 2017. The data consists of nine columns categorized by cocoa bean types for various regions and contains cocoa bean origins, flavor ratings, manufacturers name, reference numbers, % of cocoa, and the year of the review.

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| Company (Manufacturer) | Name of company that produced the chocolate bar |
| Specific Bean Origin or Bar Name | Name of the chocolate bar |
| REF | Reference number that links to the rating source |
| Review Date | Date the chocolate was reviewed |
| Cocoa Percent | Percentage of cocoa in the chocolate bar |
| Company Location | Location of the company that produced the chocolate bar |
| Rating | Expert rating of the chocolate bar on a scale of 1 to 5 |
| Bean Type | Chocolate bean used (variety) |
| Broad Bean Origin | Region where beans were grown |

**Data Limitations:**

The data was collected from a Public Doman with open licenses to allow creators to share their work, which presents an opportunity for bias and data inaccuracies.

**Data Relevance:**

This data set meets the data requirements for this project, as it is open source, contains a geospatial component, and it meets the size and variable requirements. This project is older than three years old but was provided by Career Foundry as an available dataset that meets the specific criteria for this achievement.

**Cleaning Data**

A screenshot of a computer

Description automatically generated

**Understanding Data**

A screenshot of a computer

Description automatically generated

**Key Questions**

* **What countries produce the highest rated chocolate bars?**
* **Where are the best cocoa beans grown? Specifically, do chocolate bars with the highest % of cocoa have the highest ratings?**
* **What are the top five countries with the highest ratings?**
* **What are the top five companies that have the highest ratings, and over what period?**